

AI & forskningsdataprocessen

SND:s nätverksträff 2025-10-23

Karin W. Tikkanen



April: 2025 GAIA Conference (Göteborg)

- Många företagsrepresentanter: SAAB, Siemens, AstraZeneca, AI Sweden etc.
- Huvudtemat: "allting man kan göra med hjälp av AI (så länge lagen inte sätter stopp"



Keynote av professor Kristina Höök, KTH

"Intelligence is Corporeal — Rethinking Design Through the Body"

Alla presentationer som gavs på huvudscenen hittas på:
https://www.youtube.com/@GAIA_GBG



Juni: Open Repositories 2025 (Chicago)

- Forskarstödjande personal samt representanter från olika repositorer/bibliotek
- >400 deltagare från hela världen
- Huvudfrågor: samarbete som styrka, öppen vetenskap som ett redskap för att stärka demokratin.
- AI nämndes i en majoritet av papers.
- "Few of us really understand how they [AI programs] work. Therefore we tend to inscribe intelligence on them."/Ben Zhao



Valda citat: positivt om AI

- "Repositories need to be made future safe, adaptable to AI."
- "Repositories can function as collaborative assistants. AI agents can reduce redundancy and boost interoperability. Questions you can ask: Is this dataset complete? Was there any bias? Did they achieve what they set out to do?"
- "We tried using AI to fill out metadata schemes. It needs fine-tuning, but about 80% turned out more or less correct."



Valda citat: negativt om AI

- Om *oral history* och *video history collections*: Personer som gav tillåtelse till publicering gjorde det *före* AI. Nu kan AI-agenter ladda ner bild och ljud för att använda i återskapande av röster och bilder etc.
- Om AI och crawlers: En sökträff ger ett svar samt data, man behöver inte längre gå till repositoret för att aktivt ladda ned data. Repositoret riskerar att bli osynligt. Man lägger ned tid på forskningsdatahantering, men får inget erkännande.
- "80% of web traffic is bots/artificial traffic. There are more bots online than people! AI scrapers are draining websites, raising costs."

The Usual Caveat First...

Generative AI is fundamentally different from “discriminatory” AI, aka deep neural networks used for classification.

Those “good” AI systems power most advances in drug discovery, self-driving cars, medical imaging, climate modeling, recommendation systems ...

GenAI Hype Cycle Past Its Peak ...

GS Goldman Sachs

Gen AI: too much spend, too little progress

The promise of generative AI technology to transform societies is leading tech giants and beyond to invest heavily in the technology.

Ju CNN

Has the AI bubble burst? Wall Street says generative intelligence will never make money

There's been one big question on the minds of Wall Street investors: When will anyone start making actual money from AI? The answer, according to a new report from Goldman Sachs, is "never."

1 month ago

F Futurism

<https://futurism.com/the-byte/google-researchers-paper...>

Google Researchers Publish Paper About How AI Is ...

Jul 4, 2024 — Google researchers have come out with a new paper that warns that **generative AI is ruining** vast swaths of the **internet** with fake content.

CNBC

Generative AI requires massive amounts of power and water, and the aging U.S. grid can't handle the load

Data centers are being built at a rapid pace to support generative AI, and concerns are mounting about whether we can generate enough power to support them.

Jul 28, 2024

B Barron's

AI is hitting a wall just as the hype and speculation reaches the stratosphere

1 day ago



Analysis by Allison Morrow, CNN

3 minute read · Published 5:30 AM EST, Tue November 19, 2024

Inc.

NEWSLETTERS

SUBSCRIBE

SEARCH

AI Hits a Wall: What This Means for the Tech Industry

AI commercialization is a game of musical chairs, and the music has stopped. For now.

EXPERT OPINION BY JOE PROCOPIO, FOUNDER, JOEPROCOPIO.COM @JPROCO

DEC 5, 2024

Chain of Thought Reasoning (Google Deepmind 2022)

- Prompt LLM to decompose problem into intermediate steps
- Then solve each problem in turn
- Benefits
 - Less likely to make errors on smaller tasks
 - More interpretable

LLM can give steps and walk user through steps (even more convincing!)

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓



Oktober: Open Science 2025 (Hamburg)

- Huvudtemat: hur AI påverkar synen på, och kan samverka med, öppen vetenskap.
- Konferensen riktade sig i huvudsak till forskare. Ett 40-tal var med på plats, ca 100 med online.
- Om vikten av att vara kritisk till all forskningsoutput som tas fram med hjälp av AI-redskap.
- Är det dags för en uppdatering av FAIR, till FAIR-**R**?
 - **R = Responsibly Licensed**
 - "AI requires datasets that are both technically and legally reusable."

45

In your own research or practice, have you noticed that openness is becoming harder to pursue — or even coming under attack — in the age of AI? For example, via infrastructural strain, loss of context in pursuing better machine-readability, lack of recognition for opening knowledge, challenges in governing data reuse, or the fragility of reproducibility in AI/ML workflows?





Artikeltips

- Hosseini et al. (2024) "Open Science at the generative AI turn: An exploratory analysis of challenges and opportunities", *Quantitative Science Studies* 6, 22–45.





Hur man pratar om AI på konferenser (2025)

1. Vilka allmänna begränsningar vill vi ska gälla för användningen av AI framöver?
2. Vad händer med synligheten (för forskaren/repositoriet) när crawlers gör att användare inte längre behöver gå till den ursprungliga källan för att hitta information?
3. Vad gör punkt 2 med upphovsrätten och referenshanteringen?
4. Vad gör punkt 2 och 3 med forskares vilja att bidra till öppen vetenskap?
5. Vad innebär användningen av AI för klimatet?

Fortsättning följer...

Karin.tikkanen@snd.se

